**ANALYZING PRICING, REVIEWS, AND STAR RATING PREDICTABILITY IN THE AIRBNB MARKET**

**Table of Contents**

[Introduction 3](#_Toc161411764)

[Data Description 3](#_Toc161411765)

[Research Questions 3](#_Toc161411766)

[Methodology 3](#_Toc161411767)

[Data Processing Details 4](#_Toc161411768)

[Statistical or Machine Learning Approach 5](#_Toc161411769)

[Results and Analysis 6](#_Toc161411770)

[Conclusions 11](#_Toc161411771)

[Recommendations 11](#_Toc161411772)

[References 13](#_Toc161411773)

# Introduction

The Airbnb service dynamics is the focus of this study which is examined via the combined techniques of SQL queries and machine(s) learning models. The study starts by trying how the average price of listings varies depending on room type, property type, and market, which reveals how the trends of apartments and houses go in every region. This part of the survey concentrates on the average count of reviews on listing units in various markets to disclose the amount of engagement and popularity of the listings in different areas. Besides that, the study is focused on the question of the predictability of the star rating for listings based on the provided attributes such as room type, property type, price, and market, using regression models to estimate the predictive accuracy.

# Data Description

The dataset applied for the study consists of data linked to the Airbnb listings which includes property characteristics such as room, property type, price, market company star rating and rank. This approach is an all-in-one solution to obtain aggregated Airbnb rentals in various markets and allows a comprehensive look at pricing trends, popularity indicators (rating, number of reviews), and the relation between attribute listing and star rating. The preprocessing thus involves eliminating the missing values which maintain the set’s integrity. These data are crucial in building the background for investigating how the Airbnb listings’ prices and popularity are affected by the market factors in the Airbnb marketplace

# Research Questions

Q1. How does the average price of listings vary between different room types and property types in each market?

Q2. What is the average number of reviews per listing in each market?

Q3. How accurately can predict the star rating of a listing based on its features such as room type, property type, price, and market?

# Methodology

The method consists of three main steps which include cleaning the data, using SQL queries to look at the data to answer specific research questions, and using machine learning models to guess the star scores. The process of data cleaning has been used to ready the data for the purpose of the analysis (Zolbanin and Wynn, 2023). SQL queries are used to find the average price and the number of ads in each area. In order to predict star scores based on offering details, machine learning models are created and tested. Random Forest, Linear Regression and gradient boosting regression models have been used as an example of this machine learning analysis.

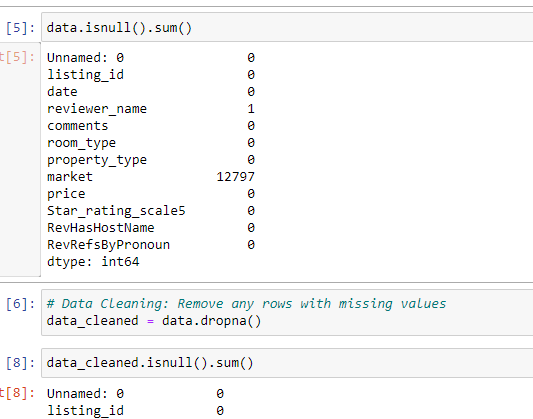


**Figure 1: Importing libraries and dataset**

(Source: Retrieved from Jupyter Notebook)

The above figure shows that several libraries have been imported for the purpose of the analysis where the pandas libraries have been used to load the dataset. The dataset has been loaded with the help of its read\_csv() method.

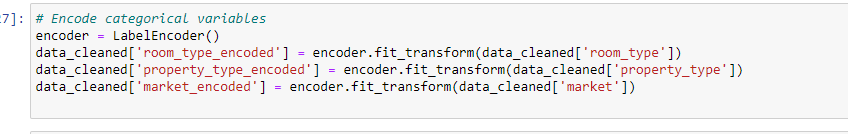
## Data Processing Details



**Figure 2: Checking null values**

(Source: Retrieved from Jupyter Notebook)

In this phase, the null values from the dataset have been checked with the help of the isnull().sum() method which shows that the column reviewer\_name and the market columns have the null values. The rows which have null values have been removed with the help of the dropna() method.

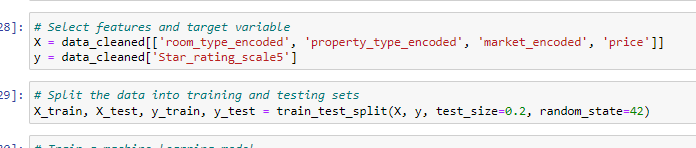


**Figure 3: Label encoding**

(Source: Retrieved from Jupyter Notebook)

In this step, group factors like market, property type, and room type are turned into numbers using the "LabelEncoder" function from scikit-learn. So that the machine learning models can understand them correctly, these variables have unique numbers assigned to each group. The different types of rooms, like "Private room" and "Entire home/apt," are turned into numbers. Due to this change, categorical data can now be used as features in machine learning models.

## Statistical or Machine Learning Approach

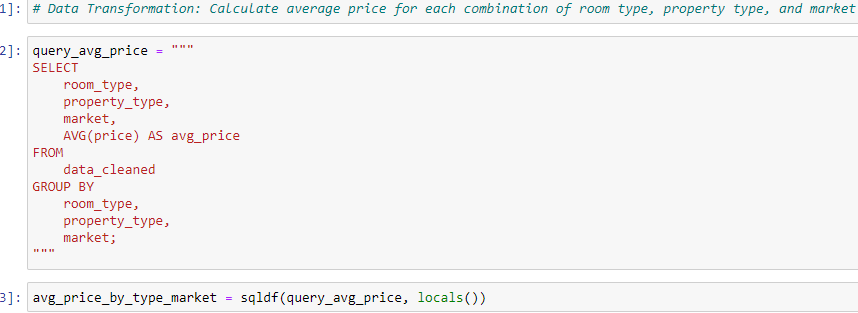


**Figure 4: Data Preparation for Machine Learning Model**

(Source: Retrieved from Jupyter Notebook)

Before performing any kind of machine learning model analysis, data preparation is an important stage. The feature as well as the target variable have been chosen based on the research question. The data has been split in 80 and 20 percent ratio which is the main stage of the analysis. The statistical or machine learning approach is to show predictive models how to understand the links between factors (like room type, property type, price, and market) and the outcome variable (like star rating). Gradient boosting, random forest, and linear regression are a few of the regression strategies utilized in this study to figure out the star review. In arrange to assess how well the models guess star grades, the study has utilized measurements such as mean squared error (MSE) and R-squared scores.

# Results and Analysis



**Figure 5: Query to calculate average price**

(Source: Retrieved from Jupyter Notebook)

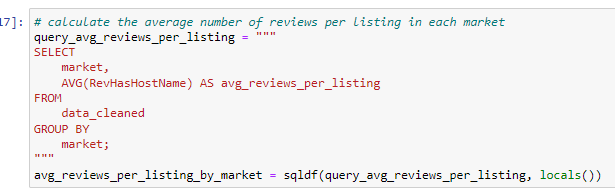
The query figures out what the average cost of advertisements is based on different sorts of rooms, properties, and areas. The information is sorted out into groups that show how costs vary between those groups. The avg\_price\_by\_type\_market dataset, which contains the average cost for each room sort, property sort, and advertising mix, makes a difference in seeing at pricing patterns within the Airbnb postings dataset in more detail and see how they change over time.



**Figure 6: Result of the query**

(Source: Retrieved from Jupyter Notebook)

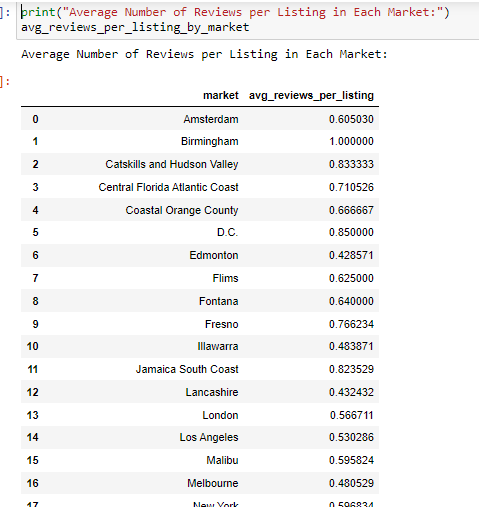
This report shows the average cost of advertisements broken down by area, property sort, and room sort. As a case, it appears that the average cost of a house or flat in Amsterdam is nearly $151.25. The table shows the by and large average costs for a number of distinctive sorts of properties and gives valuable details about how costs are changing in different ranges and sorts of properties.



**Figure 7: Query to calculate the average number of reviews**

(Source: Retrieved from Jupyter Notebook)

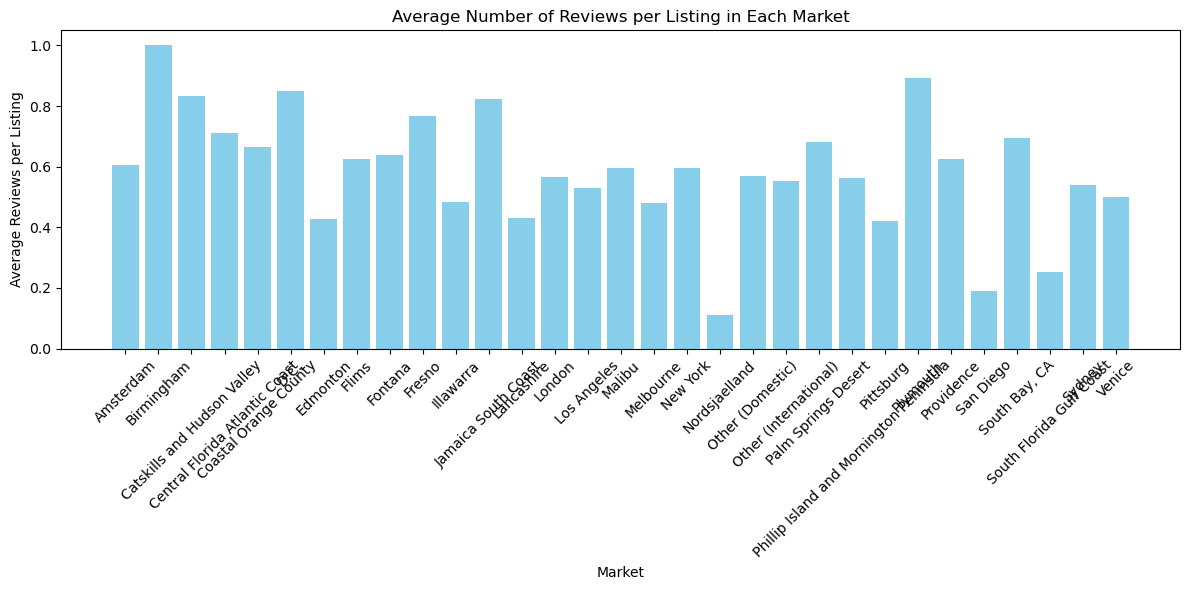
Each market's code tells the average sum of reviews for each survey. A field called "RevHasHostName" is used to show how many reviews an item has gotten. Showing the average number of comments or interactions per post in each market, the result gives information about review activity in several areas. Such details can help to figure out how famous or happy people are with commercials in various locations.



**Figure 8: Result of the query**

(Source: Retrieved from Jupyter Notebook)

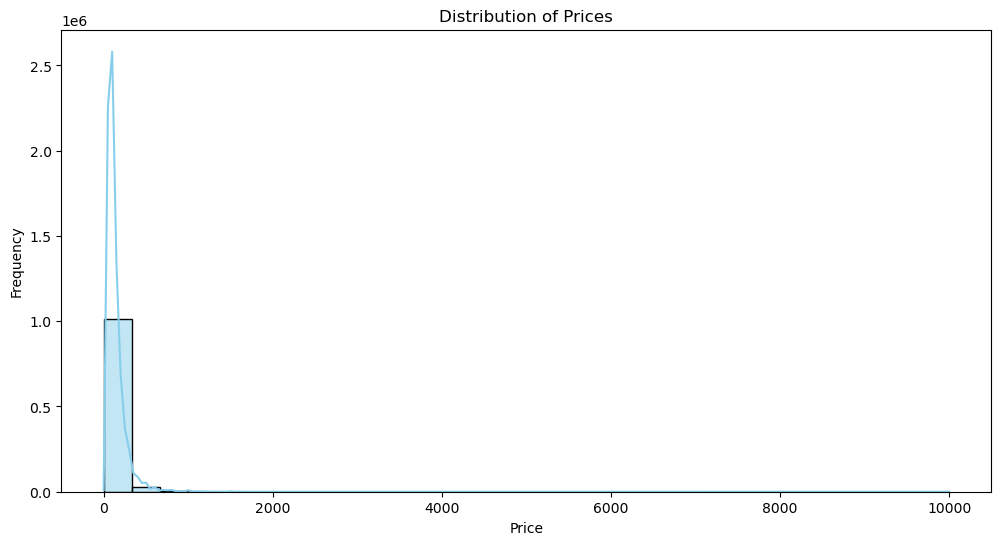
The result shows the average amount of reviews for each listing in each area. This data shows how many people responded to or interacted with posts in different areas. People may like or be more interested in markets that have more average reviews per listing. On the other hand, markets with fewer average reviews per ad may have less traffic or even less pleasant visitors.



**Figure 9: Bar plot for average number of reviews per listing**

(Source: Retrieved from Jupyter Notebook)

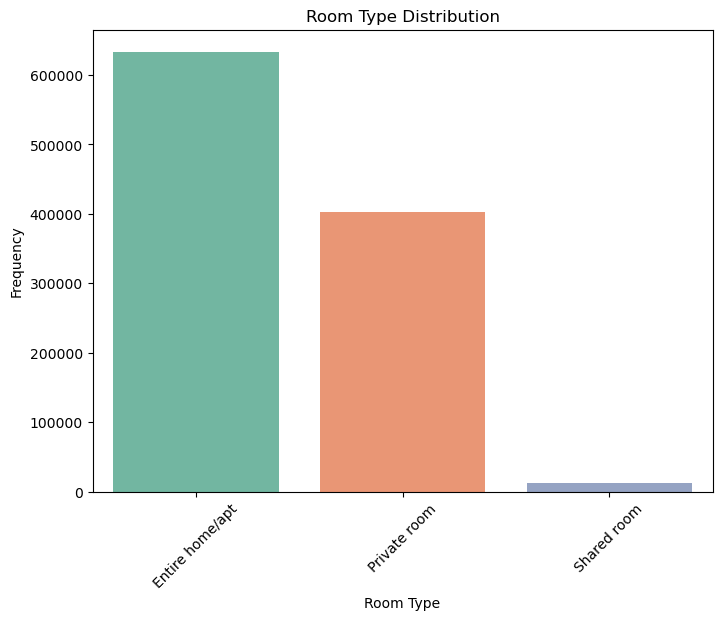
The above figure shows that a bar plot has been used to find the average number of reviews per listing in each market with the help of the query result. The result shows that Birmingham has highest number of reviews per listing.



**Figure 10: Distribution of prices plot**

(Source: Retrieved from Jupyter Notebook)

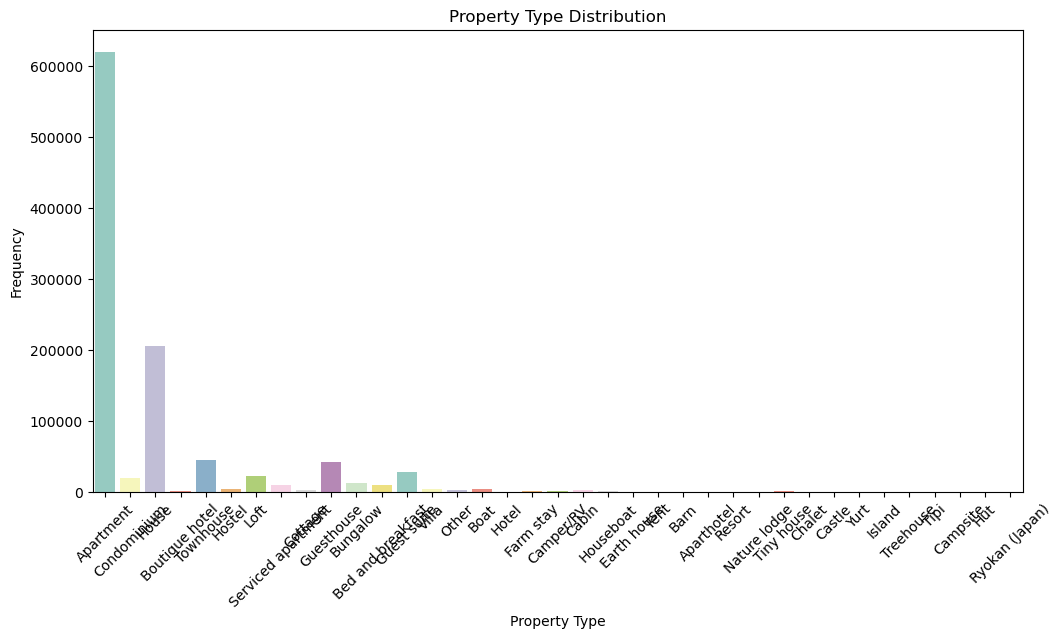
The graph shows how the prices in the dataset are spread out. Most of the prices are in the middle, which could mean that the spread is normal. This graph shows the range of prices from 0 to 10,000 on the x-axis and the number of times each price appears on the y-axis. The peak in the middle shows that prices are more normal in the middle range, since there are fewer ads at the very low and very high price points.



**Figure 11: Room type distribution plot**

(Source: Retrieved from Jupyter Notebook)

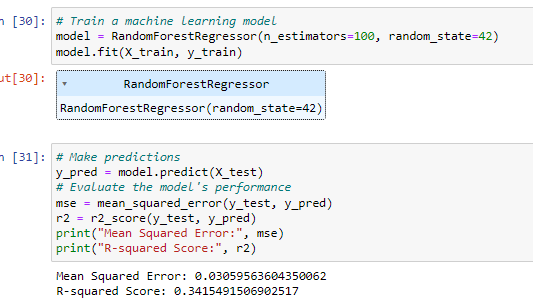
The above figure shows that the distribution of the room type has been visualized in the part where it can be seen that most types of room features in the dataset are Entire homes/apt.



**Figure 12: Property type distribution plot**

(Source: Retrieved from Jupyter Notebook)

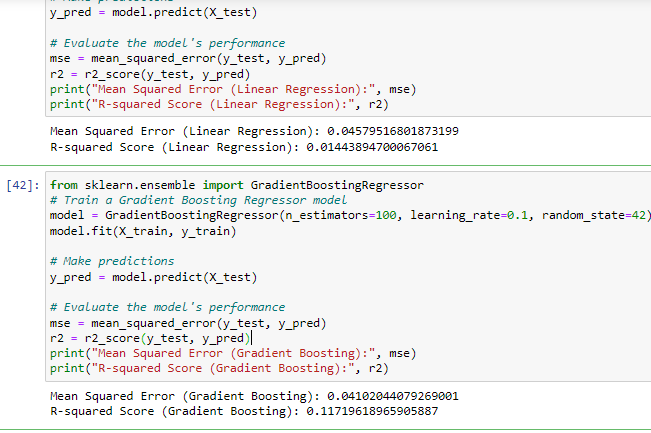
The distribution of property types has been shown here with the help of the count plot where it has been observed that the feature apartment has the most occurrences in the dataset.



**Figure 13: Training the Random forest regression model**

(Source: Retrieved from Jupyter Notebook)

In response to the third study question, which is about figuring out how to rate items with stars based on things like room type, property type, price, and market, a Random Forest Regressor model was trained and tested. The model had a Mean Squared Error of 0.0306 and an R-squared Score of 0.3415. As it can be seen, the model is pretty good at figuring out star scores based on the information given and it can explain about 34% of the differences between rates.



**Figure 14: Linear and Gradient Boosting regression**

(Source: Retrieved from Jupyter Notebook)

An R-squared Score of 0.0144 and a Mean Squared Error of 0.0458 were found for Linear Regression in answer to the third study question. However, Gradient Boosting gave an R-squared Score of 0.1172 and a Mean Squared Error of 0.0410. The fact that Gradient Boosting did better than the other two models suggests that it can be used to predict star ratings in this case.

# Conclusions

In conclusion, this study looked into how prices change, how consistent star rates are, and how reviews are written in the Airbnb market. It has been observed from the results that the prices have varied in the factors like areas, types as well as rooms types. Changes in the average number of reviews per ad from one market to the next also show changes in how happy and popular visitors are. The machine learning models used to guess star scores were not all accurate. It has been seen that the model gradient boosting has performed better than the others. This means that factors such as the type of room, the type of property, the price, and the market can all be used in understanding the number of stars in the listing of several Airbnb.

# Recommendations

After studying how Airbnb prices change, reviewing trends, and how stable star ratings are, some of the recommendations for hosts to improve their services and make guests happier have been described (Applin, 2023).

***Tailor Pricing Strategies***

Hosts should change how they set their prices based on market trends, the type of property, and the type of housing. The hosts of the properties can change their costs to make the most of the money when it can be seen from the study that the types of properties sell for more in several kinds of places.

***Focus on Guest Satisfaction***

Each marketplace has a different number of reviews per offering, so hosts should focus on making sure visitors have a great time to get more positive reviews (Culotta *et al.* 2023). It can be meant to make all the services available as well as keep the rooms clean by satisfying the customers.

***Optimize Listing Features***

One way for hosts to improve their listings is to highlight the traits that guests in their target market value the most. The most attractive properties can be highlighted in the listings to attract more people through it. For example, if data shows that these types of properties are more popular in a certain area, hosts may do this.

***Utilize Machine Learning for Pricing and Reviews***

Machine learning models can help hosts find patterns in their market and predict how prices will change (Contu *et al.* 2023). In order to make smart decisions as well as making decisions on pricing and connection with customers, the hosts can look at previous data as well as market trends.

# References

Applin, R., 2023. Prices, Quality, and Learning in Airbnb Markets. Available at SSRN 4409993.

Contu, G., Frigau, L. and Conversano, C., 2023. Price indicators for Airbnb accommodations. Quality & Quantity, 57(5), pp.4779-4802.

Culotta, A., Jin, G.Z., Sun, Y. and Wagman, L., 2023. Safety Reviews on Airbnb: An Information Tale (No. w31855). National Bureau of Economic Research.

Zolbanin, H.M. and Wynn, D., 2023. From star rating to sentiment rating: using textual content of online reviews to develop more effective reputation systems for peer-to-peer accommodation platforms. Journal of Business Analytics, 6(2), pp.127-139.